Preventive Medicine Reports 7 (2017) 116–118

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: http://ees.elsevier.com/pmedr

Short Communication

Comparing geographic area-based and classical population-based incidence and prevalence rates, and their confidence intervals

Ding-Geng Chen [a,b,c]

[a] School of Social Work, University of North Carolina, Chapel Hill, NC, USA
[b] Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
[c] Department of Statistics, University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Article history:
Received 16 March 2017
Received in revised form 17 May 2017
Accepted 28 May 2017
Available online 09 June 2017

Keywords:
Incidence rate
Prevalence rate
G rates
P rates
D rates
Variance
Confidence intervals
HIV/AIDS interventions

ABSTRACT

To quantify the HIV epidemic, the classical population-based prevalence and incidence rates (P rates) are the two most commonly used measures used for policy interventions. However, these P rates ignore the heterogeneity of the size of geographic region where the population resides. It is intuitive that with the same P rates, the likelihood for HIV can be much greater to spread in a population residing in a crowed small urban area than the same number of population residing in a large rural area. With this limitation, Chen and Wang (2017) proposed the geographic area-based rates (G rates) to complement the classical P rates. They analyzed the 2000–2012 US data on new HIV infections and persons living with HIV and found, as compared with other methods, using G rates enables researchers to more quickly detect increases in HIV rates. This capacity to reveal increasing rates in a more efficient and timely manner is a crucial methodological contribution to HIV research. To enhance this newly proposed concept of G rates, this article presents a discussion of 3 areas for further development of this important concept: (1) analysis of global HIV epidemic data using the newly proposed G rates to capture the changes globally; (2) development of the associated population density-based rates (D rates) to incorporate the heterogeneities from both geographical area and total population-at-risk; and (3) development of methods to calculate variances and confidence intervals for the P rates, G rates, and D rates to capture the variability of these indices.

1. Introduction

In response to the U.S. President's Emergency for AIDS Relief (PEPFAR) initiative that seeks to end the HIV/AIDS epidemic, extensive research and comprehensive assessment measures have been developed to assess the impact of the HIV/AIDS epidemic and the changing rates of HIV transmission (U.S. Department of State, 2012). These measures have been used to better inform the development of evidence-informed interventions and policy makers' decisions related to funding for HIV programs.

One of the most important measures is the classical population-based incidence rate and prevalence rate—commonly called P rates—that has been widely used as the traditional epidemiologic measure to provide quantitative data on the total number of infections, incidence, and prevalence among the total population. Specifically, the two population-based P rates, P incidence rate (*P.IR*) and P prevalence rate (*P.PR*), are defined as follows:

$$P.IR = (\text{Newly infected persons in a year})/(\text{Population-at-risk in } 100,000) \quad (1)$$

$$P.PR = (\text{All infected persons in a year})/(\text{Population-at-risk in } 100,000) \quad (2)$$

As seen from the formulae above, the P rates are scaled by the total population-at-risk, and therefore, P rates are heavily dependent on the heterogeneity of the population-at-risk. For example, as reported by the U.S. Centers for Disease Control and Prevention (CDC), the *P.PR* in the United States during 2014 was 18.5/100,000; this calculation expresses the ratio of the 1,218,000 persons living with HIV (PLWH) over the total U.S. population. However, the *P.PR* for Botswana in 2014 had a much higher value of 25,200/100,000 although the Botswana population had a much smaller number of PLWH than the United States (39,000 vs. 1,218,000 respectively). Therefore, to better reflect the impact of HIV on the total population, the P rates should be refined to account for geographic heterogeneity based on the fact that geographic area size plays a crucial role in the HIV/AIDS epidemic. It is intuitive that for the same P rates, the likelihood can be much greater for HIV to spread from one to another in a population residing in a crowed small urban area than the same number of population residing in a large rural area. Simply mapping a P rate by geographic areas does not provide complete information about geographic differences of the

E-mail address: dinchen@email.unc.edu.

http://dx.doi.org/10.1016/j.pmedr.2017.05.017

HIV/AIDS epidemic, underscoring the need for new measures (Chen and Wang, 2017).

This knowledge gap is filled by the G rates proposed by Chen and Wang (2017), which incorporate the geographical heterogeneity and define the geographic area-based incidence rate and prevalence rate (i.e., G rate) as the number of persons newly infected or living with HIV in one year within a jurisdiction or region over the total geographic area. Moreover, Chen and Wang used the defined G rates to assess incidence (i.e., G incidence rate or *G.IR*) and prevalence (i.e. G prevalence rate or *G.PR*) of HIV. Specifically,

$$G.IR = (\text{Newly infected in one year})/\left(\text{Total geographical area in 1000 km}^2\right) \tag{3}$$

$$G.PR = (\text{Total infected in one year})/\left(\text{Total geographical area in 1000 km}^2\right) \tag{4}$$

Ultimately, these G rates measure the geographic density of a disease. In addition to using the G rates to quantify the HIV/AIDS epidemic, the G rates can be used to analyze incidence and prevalence of many other diseases, thus informing public health planning and decision making. G rates utilize the total geographic area of a region, thereby enhancing the classical P rates with the total population-at-risk as the reference. However, we caution the researchers that the use of geographic incidence/prevalence rates should limited to homogenous areas.

In their paper, Chen and Wang effectively demonstrated the applicability of their proposed G rates using data on new HIV infections and persons living with HIV in the United States during 2000–2012. The authors used the G rates to evaluate rates of HIV infection across individual U.S. states and in relation to the geographic size of each state. Based on their analyses of the U.S. sample, Chen and Wang concluded that the G rates time trends revealed increases in HIV transmission more quickly than did the P rates. The ability to detect changes and trends earlier in time has important implications for informing the design and implementation of strategically targeted, precision interventions needed to achieve the goals of the AIDS-Free Generation initiative (U.S. Department of State, 2012; UNAIDS, 2010). The cross-plot of P rates with G rates in Fig. 3 from Chen and Wang (2017) is a masterpiece!

2. Comments and discussions

To enhance the newly proposed G rates, I offer three suggestions for further development of this important new measure.

2.1. Analyze global HIV epidemic data using G rates

As demonstrated in Chen and Wang (2017), G rate analysis has the capacity to capture geographical heterogeneity that conventional P rates cannot. Although Chen and Wang presented an analysis of U.S. data, those data are more homogenous than worldwide data. Therefore, analysis of worldwide HIV data would provide an even better demonstration of the superior capacity of G rates. In fact, global HIV epidemic data are available from UNAIDS (http://aidsinfo.unaids.org/).

2.2. Fully develop the D rates

Chen and Wang (2017) briefly discussed the concept of D rates that would synchronize and standardize the effects from both the total population-at-risk (i.e., the effects from the P rates) and the total geographic area at risk (i.e., the effects from the newly proposed G rates). Given that the D rates would incorporate the heterogeneities from both geographical area and total population-at-risk, I recommend defining the D incidence rate (*D.IR*) and D prevalence rate (*D.PR*), as follows:

$$D.IR = (\text{Newly infected in one year})/(\text{population density})$$

$$D.PR = (\text{Total infected in one year})/(\text{Population density})$$

These D rates can be represented by the G rates and P rates as follows:

$$\begin{aligned} D.IR &= (\text{Newly infected in one year})/(\text{Total population-at-risk}/\text{Total geographical area}) \\ &= P.IR \times (\text{Total geographical area}) \\ &= G.IR \times (\text{Total geographical area})^2/(\text{Total population-at-risk}) \end{aligned} \tag{5}$$

$$\begin{aligned} D.PR &= (\text{Total infected in one year})/(\text{Total population-at-risk}/\text{Total geographical area}) \\ &= P.PR \times (\text{Total geographical area}) \\ &= G.PR \times (\text{Total geographical area})^2/(\text{Total population-at-risk}) \end{aligned} \tag{6}$$

2.3. Develop statistical confidence intervals for the P rates, G rates, and D rates

To capture the statistical variability of HIV/AIDS data with measurement errors, researchers will need measures of variance and confidence intervals (CIs) for these rates. This enhancement can be accomplished based on the standard statistical theory. In fact, P rates can be treated as binomial-distributed data and the G rates as Poisson distributed data (Casella and Berger, 2002; Chen et al., 2017), and D rates are derived from the two.

To illustrate the implementation, formulae are developed and presented below for calculating the variances and CIs of *P.IR*, *G.IR* and *D.IR*. These formulae can be similarly adapted to estimate variance and CIs for *P.PR*, *G.PR*, and *D.PR*.

2.3.1. Variance and 95% CI for P incidence rate

For the P incidence rate (i.e., *P.IR*), the variance can be estimated as follows:

$$var(P.IR) = \frac{P.IR \times (1 - P.IR)}{\text{Population at Risk}}. \tag{7}$$

From the estimated variance in Eq. (7), the 95% CI can be constructed as $(P.IR - 1.96 \times \sqrt{var(P.IR)},\ P.IR + 1.96 \times \sqrt{var(P.IR)})$.

2.3.2. Variance and 95% CI for D incidence rate

Since the D incidence rate (i.e., *D.IR*) can be directly calculated from the P incidence rate, as seen in Eq. (5), its variance can then be formulated as $\text{var}(D.IR) = \text{var}(P.IR) \times \text{Total Geographical Area}^2$; the associated 95% CI would be $(D.IR - 1.96 \times \sqrt{var(D.IR)},\ D.IR + 1.96 \times \sqrt{var(D.IR)})$.

2.3.3. Variance and 95% CI for G incidence rate (G.IR)

The *G.IR* is calculated using the data on the new cases of infection from a given geographical area. If the number of newly infected cases is regarded as Poisson counts, the variance of *G.IR* can be formulated based on the statistical property of Poisson distribution where the variance equals the mean, that is,

$$var(G.IR) = \frac{\text{Newly Infected Cases}}{\text{Total Geographical Area}^2} = \frac{G.IR}{\text{Total Geographical Area}}. \tag{8}$$

Consequently, the 95% CI for *G.IR* can be constructed using the estimated variance in Eq. (8), which is $(G.IR - 1.96 \times \sqrt{var(G.IR)},\ G.IR + 1.96 \times \sqrt{var(G.IR)})$.

2.4. Numerical illustrations

For illustration, consider the data from Alabama as reported in Table 1 from Chen and Wang (2017). The total geographical area of Alabama is 131,171 km^2, the total population-at-risk is 4,022,346, and the *G.IR* is 4.88/1000 km^2. Based on Eq. (3), the newly infected HIV cases would be $4.88 \times 131{,}171 = 640$; this number (640) will be used for calculating variances and CIs.

2.4.1. Calculations for P incidence rate (P.IR)

Again using the Alabama data, we can calculate the *P.IR* using Eq. (1), which is $P.IR = 640/4{,}022{,}346 \times 1000 = 0.159/1000$ population; this calculation reproduces the value in Table 1 of Chen and Wang (2017). Its variance can be calculated using Eq. (7):

$$Var(P.IR) = \frac{0.1591 \times (1-0.1591)}{4,022,346} \times 1,000 = 3.326 \times 10^{-5}. \quad (9)$$

In the above calculation (Eq. 9), the *P.IR* is scaled within 0 and 1 so that Eq. (7) can be applied. With this estimated variance, the 95% CI can be calculated as $(P.IR - 1.96 \times \sqrt{var(P.IR)},\ P.IR + 1.96 \times \sqrt{var(P.IR)}) = (0.1591 - 1.96 \times \sqrt{3.326 \times 10^{-5}},\ 0.1591 + 1.96 \times \sqrt{3.326 \times 10^{-5}}) = (0.1478, 0.1704)$. We can conclude that the *P.IR* rate for Alabama in 2012 was 15.91 with 95% CI [14.78, 17.04] per 100,000 people.

2.4.2. Calculations for D incidence rate (D.IR)

Since the *D.IR* is directly related to the *P.IR* as defined in Eq. (5), we can make use of the calculations above for the *P.IR* to obtain the estimated *D.IR*, its variance, and the associated 95% CI, that is, $D.IR = 0.1591 \times \frac{131{,}171}{100{,}000} = 0.2087$ (unit of 100 people per km2) and its $var(D.IR) = 3.326 \times 10^{-5} \times (\frac{131{,}171}{100{,}000})^2 = 5.723 \times 10^{-5}$; the associated 95% CI can be calculated as $(D.IR - 1.96 \times \sqrt{var(D.IR)},\ D.IR + 1.96 \times \sqrt{var(D.IR)}) = (0.2087 - 1.96 \times \sqrt{5.723 \times 10^{-5}},\ 0.2087 + 1.96 \times \sqrt{5.723 \times 10^{-5}}) = (0.1939, 0.2235)$.

2.4.3. Calculations for G incidence rate (G.IR)

The estimated $G.IR = \frac{640}{131{,}171} \times 100 = 0.4879$ /100 km^2 and the variance can be calculated based on Eq. (8) as $var(G.IR) = \frac{640}{(131{,}171/100)^2} = 3.7197 \times 10^{-4}$. Consequently the 95% CI for *G.IR* can be obtained as $(G.IR - 1.96 \times \sqrt{var(G.IR)},\ G.IR + 1.96 \times \sqrt{var(G.IR)}) = (0.4879 - 1.96 \times \sqrt{3.7197 \times 10^{-4}}, 0.4879 + 1.96 \times \sqrt{3.7197 \times 10^{-4}}) = (0.4501, 0.5257)$.

3. Conclusion

The G rates proposed by Chen and Wang (2017) are an important contribution to public health research and fill a critical knowledge gap by capturing geographic heterogeneity. The enhancements to the G rates proposed in this article will allow researchers to fully exploit the superior capacity of the G rates to detect changes in infection rates of a disease, thus better informing the decision making of public health officials.

Acknowledgement

I would like to thank the two anonymous reviewers and Professor Eduardo L. Franco for their comments which substantially improved the quality of this article.

References

Casella, G., Berger, R.L., 2002. Statistical Inference. 2nd ed. Duxbury Press, Pacific Grove, CA.

Chen, X., Wang, K., 2017 March. Geographic area-based rate as a novel indicator to enhance research and precision intervention for more effective HIV/AIDS control. Prev. Med. Rep. 5:301–307. http://dx.doi.org/10.1016/j.pmedr.2017.01.009.

Chen, D.G., Peace, K.E., Zhang, P., 2017. Clinical Trial Data Analysis Using R and SAS. Chapman and Hall/CRC Press, Boca Raton, FL.

U.S. Department of State, 2012. PEPFAR Blueprint: Creating an AIDS-Free Generation. Office of the U.S. Global AIDS Coordinator and Health Diplomacy, Washington DC Available at:. https://www.pepfar.gov/documents/organization/201386.pdf.

UNAIDS, 2010. On the Fast Track to End AIDS: 2016–2021 Strategy. Geneva, Switzerland. Available at:. http://www.unaids.org/sites/d1efault/files/media_asset/20151027_UNAIDS_PCB37_15_18_EN_rev1.pdf.